GeoAI for Cities: Building Class Identification via Neural Network-Based Analysis of Satellite Imagery

*Mohammed Sohaib Uddin1, Madhu Goutham Reddy Ambati2,Abdul Rahman Shaikh2,\*, M. Courtney Hughes3, Mahdi Vaezi4*

***1****College of Business, Department of Operations Management and Information Systems, Northern Illinois University, Dekalb, IL*

***2****College of Liberal Arts and Sciences, Department of Computer Science, Northern Illinois University, Dekalb, IL*

***3****College of Health and Human Sciences, School of Health Studies, Northern Illinois University, Dekalb, IL*

***4****College of Engineering and Engineering Technology, Department of Engineering Technology, Northern Illinois University, Dekalb, IL*

*\*Corresponding author: Abdul Rahman Shaikh, College of Liberal Arts and Sciences, Department of Computer Science, Northern Illinois University, Dekalb, IL,* [*Z1841128@students.niu.edu*](mailto:Z1841128@students.niu.edu)

**Abstract**

Urban expansion demands precise classification of diverse building types to support sustainable urban planning, efficient resource allocation, and effective disaster response. However, irregular building orientations, pervasive shadows, and extensive geographic variability continue to challenge standard classification methods. In this study, we employ DenseNet-201, a deep learning architecture celebrated for its dense connectivity and effective feature reuse, to classify buildings into seven distinct categories. We leverage a newly curated dataset of 15,329 high-resolution (512×512 pixel) satellite images spanning all 50 U.S. states, ensuring comprehensive geographic representation and capturing wide-ranging urban conditions. Our DenseNet-201 model achieves an overall test accuracy of 84.39%, outperforming alternative architectures while displaying notable robustness against distortions arising from shadows and variable building orientations. Furthermore, we integrate a novel segmentation module using ReFineNet, coupled with test-time augmentation (TTA) and morphological watershed post-processing, to enhance the separation of closely spaced buildings and reduce classification ambiguity. A user-friendly interface allows customization of the model for specific regional requirements and evolving urban layouts. Our approach thus advances automated building classification by blending cutting-edge deep learning architectures with rigorous data preparation and geospatial analytics.

**1. Introduction**

Buildings form the backbone of urban landscapes, shaping skylines and influencing economic, social, and infrastructural development. In recent years, advances in satellite imagery, artificial intelligence (AI), and computer vision (CV) have revolutionized our ability to segment and classify these structures, enabling applications in urban planning, disaster response, energy management, and land-use analysis (Farajzadeh et al., 2023; H. Wu et al., 2020). Moreover, integrating detailed building segmentation with census data has opened new avenues for refined population estimation at county and state levels, offering valuable insights for policymakers and urban planners alike. Several studies have introduced diverse datasets and corresponding models for building segmentation and classification across various regions, highlighting the potential and challenges of these approaches (Arruda et al., 2024; Huang et al., 2022). Recent advancements in deep learning architectures, ranging from convolutional neural networks (CNNs) for object detection to transformer-based models for semantic segmentation, have demonstrated significant potential in automating large-scale building classification. However, limitations in existing labeled datasets and inconsistencies in data quality across different geographic regions continue to pose challenges for achieving reliable generalization and operational deployment.

Many existing studies focus on classifying only two or three building types, which does not reflect the rich diversity of real cities (Alsabhan et al., 2022; Vasavi et al., 2023). Accurate boundary detection remains especially challenging in dense, shadowed, or region-specific urban environments (Huang et al., 2022; Lloyd et al., 2020). Additionally, researchers lack access to comprehensive datasets that include a broad array of building types from varied cultural and architectural contexts. This constraint hinders the development of universally robust methods (Dimassi et al., 2021; Ji et al., 2019). Building classification typically involves categorizing structures based on their physical attributes, spatial distribution, and contextual surroundings (Ithape et al., 2023). Larger, well-structured datasets are therefore critical for advancing the reliability and global applicability of building classification.

The task of accurately classifying buildings presents many interconnected challenges in today’s rapidly evolving urban landscapes. Cities are dynamic, with buildings coming in countless shapes, sizes, and uses, often changing their purposes over time as communities grow and adapt (Ithape et al., 2023). Modern buildings frequently serve multiple functions, such as apartment complexes with ground-floor retail spaces or office buildings that include both commercial and residential areas, making them particularly difficult to categorize. Traditional field surveys where people walk through cities to document structures can provide detailed information but are increasingly impractical, requiring significant time and resources that quickly become outdated (Adha et al., 2022; Hu et al., 2023). This is where satellite technology has transformed our approach to building classification.

Modern satellites can capture detailed images of entire cities in minutes, providing regular updates that reveal new construction, demolition, and changes in land-use patterns (Abburu & Babu Golla, 2015; Vasavi et al., 2023). These high-resolution images offer detailed insights into building layouts with minimal effort compared to traditional surveying, making them indispensable for monitoring urban development (Alsabhan et al., 2022). However, traditional approaches often struggle to differentiate buildings in dense urban settings, where architectural similarities, shadows, occlusions, and overlapping structures obscure key visual features (Alsabhan et al., 2022; Erdem & Avdan, 2020; Vasavi et al., 2023). Trees can overlap or cast shadows on buildings, roads may look similar to rooftops, and natural features like water or terrain can produce reflections and shadows that obscure critical details. The challenge becomes even more complicated when considering the variety of buildings themselves from sprawling industrial complexes to tightly packed urban homes, each with unique roof designs, orientations, and architectural features (Huang et al., 2022; Ji et al., 2019). Shadows cast by tall buildings can hide smaller structures, and different times of day or seasonal changes can alter how buildings appear, demanding sophisticated methods that can reliably distinguish these variations (Lloyd et al., 2020; Reda & Kedzierski, 2020). Additionally, regional variations in building styles, materials, and construction patterns further complicate classification tasks (Atwal et al., 2022). While researchers have made significant progress in developing methods to overcome these obstacles, current approaches still face important limitations.

These challenges highlight the necessity for advanced, scalable, and robust classification techniques to enhance the precision and reliability of urban analysis. Addressing these challenges and limitations, our paper introduces a high-resolution satellite imagery dataset (15,329 images) specifically designed to address gaps in existing resources. It emphasizes challenging urban scenarios, irregular architectures, occluded structures, and densely packed areas, providing a robust foundation for training models on real-world complexities. Complementing this dataset, we propose a building segmentation module that isolates building footprints from background noise. This module employs test-time augmentation techniques and advanced post-processing, including morphological operations and watershed segmentation, to separate closely attached buildings, ensuring refined region proposals for subsequent classification. The segmentation process thus contributes to reducing misclassificationsarising from overlapping structures and shadows. Building upon this segmentation framework, our classification model using DenseNet201 categorizes buildings into seven types: single-residential, multi-residential, commercial, hospital, industrial, high-rise, and school. This granularity bridges a critical gap in urban classification systems, which have traditionally oversimplified structural diversity by lumping buildings into broad categories such as residential versus non-residential (Arruda et al., 2024) .

Our research makes the following three key contributions to advance the field of building classification:

1. We present a comprehensive publicly available satellite imagery dataset of 15,329 images that capture a wide range of urban scenarios and building types, addressing the need for diverse and representative data.
2. We propose a robust model that effectively segments and classifies buildings into seven distinct categories, achieving high accuracy even in complex urban environments.
3. To enhance real-world applicability, we incorporate coordinate mapping for precise geolocation and develop an intuitive interface that allows users to adapt the system to region-specific needs, facilitating scalability across different contexts.

By addressing the limitations of current methods and providing new datasets and approaches, we aim to improve the accuracy and reliability of building classification from satellite imagery. Our work has the potential to benefit urban planners, emergency responders, and other professionals who rely on accurate building information, ultimately contributing to better-informed decisions and more efficient management of urban environments.

**2. Related Work**

Early approaches to building classification often relied on specialized datasets with annotated aerial or satellite images, typically focusing on single regions or narrow building categories. For instance, (Erdem & Avdan, 2020) used the Inria Aerial Image Labeling Dataset to classify buildings in Chicago demonstrating the potential of a modified U-Net architecture with skip connections. The model achieved 87.69% accuracy but struggled in densely constructed areas where building footprints overlapped, making it difficult to capture complex geometries fully. (Vasavi et al., 2023) expanded this to classify buildings into residential, industrial, and holy places in Nashik, India, employing a U-Net with a ResNet-34 backbone. The model achieved an accuracy of 89% yet faced challenges in distinguishing similar types in complex layouts.

CNNs have been foundational, with models like DeepLabv3+ and Attention U-Net significantly improving performance. DeepLabv3+ attained an IoU of 89–90% on the Inria dataset, outperforming earlier U-Net variants, as noted by (Ekiz & Acar, 2025). These models leverage skip connections and pyramid pooling to capture context, but their limited global receptive field can result in fragmented or false detections in complex scenes, as highlighted by (Dimassi et al., 2021). Vision transformers have recently been adopted in remote sensing to address CNNs' limitations in modeling global relationships. (Wang et al., 2024) found transformers deliver stable performance in segmentation and object detection, with early results showing improved land-use classification accuracy. A study by (Liu et al., 2020) demonstrated their effectiveness in capturing long-range dependencies, crucial for recognizing entire building structures amidst large contexts. However, transformers are computationally expensive and data-hungry, often requiring large datasets for generalization, creating a trade-off between accuracy and efficiency, as noted in a survey by (Dosovitskiy et al., 2021). (Huang et al., 2022) experimented with state-of-the-art models like Mask R-CNN and SOLOv2 on images from Beijing and Munich, though they still encountered issues with diagonal rooftops and merged building clusters.

To capitalize on both paradigms, hybrid models have emerged, typically using CNN backbones for local feature extraction and transformer modules for global context. (Chang & Zheng, 2024) introduced CTANet, combining a ConvNeXt encoder with a lightweight transformer decoder, achieving state-of-the-art F1-scores and IoUs on datasets like Massachusetts, WHU, and Inria. STransU2Net (Liu et al., 2020) integrates CNN and transformer components to extract buildings of various sizes, addressing the suboptimal performance of CNNs for larger buildings and transformers for smaller ones. Hyformer, proposed by (Dimassi et al., 2021)(Yan et al., 2023) HyFormer: Hybrid Transformer and CNN for Pixel-Level Multispectral Image Land Cover Classification, extends this approach to pixel-level classification, enhancing feature expressiveness. While these models captured both global context and fine-grained details, their focus often remained on binary building extraction rather than robust multiclass classification.

Beyond binary building detection, researchers increasingly seek finer distinctions among building types, such as single-unit homes, commercial complexes, hospitals, and high-rises. (Dimassi et al., 2021) introduced the Beirut Buildings Type Classification dataset to distinguish residential from non-residential buildings, achieving 94.8% accuracy using a RexNet model. However, many studies, like (Ji et al., 2019), focus on limited categories or regions, such as Christchurch, New Zealand, facing issues with shadows, complex backgrounds, and mixed land-use areas. (Zhao et al., 2023) proposes a framework considering internal (shape) and external (location, semantic) features, addressing multi-level interactions for residential prediction. The challenge lies in ensuring models generalize across diverse geographic contexts and architectural styles. Studies like UrbanClassifier (Fang et al., 2021) tackle this by automating typology analysis across scales, but lack nationwide or multinational datasets, hindering robustness. Architectural complexity, including varying shapes, orientations, colors, and roofing materials, further complicates efforts, with shadows and occlusions distorting visible features, as noted by (Reda & Kedzierski, 2020). They applied a Faster Edge Region Convolutional Neural Network (FER-CNN), achieving 93% accuracy across six building categories but struggling with small structures and overlapping shadows.

Early approaches frequently relied on specialized datasets, like the EuroSAT dataset by (Helber et al., 2019), leveraging CNNs such as ResNet-50 and GoogleNet to classify 27,000 images (64×64 pixels) into general land-use classes. They achieved 98.57% accuracy for broad land-cover categories but lack detailed building distinctions. (Atwal et al., 2022) incorporated OpenStreetMap (OSM) data from various U.S. counties, utilizing decision tree classifiers to categorize buildings into multiple classes. Although they reached a 98% accuracy rate, OSM's inconsistent annotations and missing attributes led to reliability issues, and the building categories remained relatively coarse. Recent efforts introduce new datasets to address these gaps. (Dimassi et al., 2021) introduced a dataset for residential/non-residential classification, while (Chen et al., 2024) created a high-resolution dataset covering urban and rural areas in China, with 170,015 buildings and over 93% accuracy using CNNs. Large-scale methods, such as large-scale individual building extraction from open-source satellite imagery (Chen et al., 2024), propose frameworks combining super-resolution and instance segmentation for wider coverage. (Kusz et al., 2021) utilized LiDAR data to classify 93,440 images from Hamilton County, Indiana, distinguishing residential from non-residential buildings using U-Net. (Hang & Cai, 2020) focused on rooftop shapes and sizes for classifying industrial and residential buildings from Gaofen-2 imagery, but their limited consideration of additional building attributes such as height or function restricted the model's real-world applicability. These advancements complement our approach, which uses a U.S.-wide dataset to ensure representation across diverse regions, climates, and architectural styles.

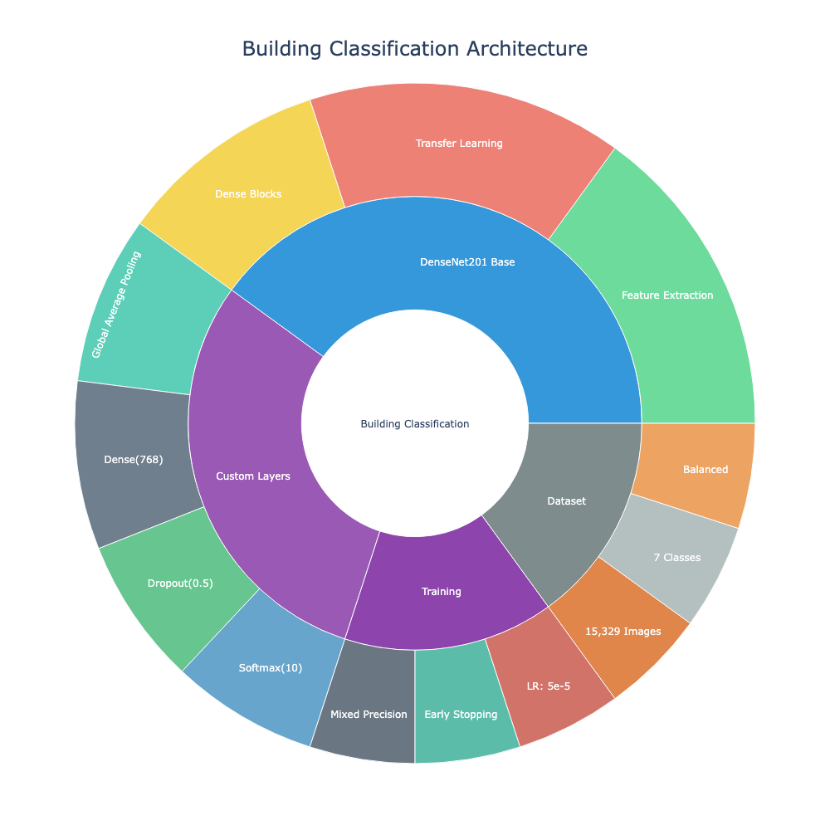
Recent advancements have further expanded the field by addressing critical segmentation challenges and generalization issues (Dabove et al., 2024; Sikdar et al., 2023). For instance, novel segmentation architectures that integrate dense skip connections and attention mechanisms (e.g., U-Net++, DeepMAO) have been developed to enhance the delineation of building boundaries, achieving superior performance in capturing complex building geometries and overlapping structures (Chen et al., 2024; Sikdar et al., 2023) . Additionally, researchers have introduced multi-scale classification techniques that leverage both local and global contextual information to improve detection accuracy across varying building sizes and densities. Recent work also highlights the importance of robust data augmentation and domain adaptation strategies to address the inherent dataset limitations and ensure models generalize well across diverse geographical regions and imaging conditions (Chen et al., 2024; Liu et al., 2020). These emerging methodologies not only refine building segmentation but also complement classification tasks by ensuring that both the identification and extraction of buildings are performed with high precision (Wang et al., 2024) .

From a remote sensing perspective, satellite imagery is not merely a passive record of the Earth's surface; it is a dynamic data source that, when analyzed with advanced machine learning models, can reveal patterns of urban development and land use. Urban morphological theory emphasizes understanding the form, structure, and evolution of cities. Within this context, building classification serves as a critical tool for interpreting how various socio-economic, environmental, and cultural factors shape the built environment. By situating our study at the intersection of these frameworks, we aim to advance the theoretical understanding of urban complexity through more accurate and detailed building classification.

A significant challenge in building classification lies in ensuring models can generalize across diverse geographical contexts and architectural styles. Our proposed research, grounded in remote sensing, urban morphology, and geospatial informatics, aims to advance building classification by integrating a DenseNet-201 architecture with a segmentation module on a comprehensive U.S.-wide dataset. This approach addresses critical gaps in existing studies, such as limited category sets and regional focus, offering deeper insights into urban complexity. Future work could explore multi-scale classification techniques and robust data augmentation strategies, as suggested by recent studies like (Chen et al., 2024), to further enhance precision and generalization.

**Table 1: Previously published building classification studies using satellite imagery**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Image Source | Year of publication | class | Study area | Channels | Image Size (Pixel) | GSD(m/pixel) | Number of Labeled Pixels (billion) | Coverage | Accuracy |
| Atwal et al. (Atwal et al., 2022) | OpenStreetMap (OSM) (*OpenStreetMap*, n.d.) | 2022 | Non-residential, residential | Fairfax County in Virginia, Mecklenburg County in North Carolina, and the City of Boulder in Colorado |  |  |  |  | United Sates | 98% |
| Li et al. (Li et al., 2021) |  | 2021 | New and old rural buildings | Rural Xinxing County, Guangdong Province, China | RGB | 900 × 900 to 1024 × 1024 pixels | 0.26 |  | China |  |
| Ji et al. (Ji et al., 2019) | Wuhan Dataset (WHU) | 2019 | countryside, residential, culture, and industrial area | Christchurch, New Zealan  d | RGB | 512 × 512 | 0.075-2.7 | 57.67 | New Zealand |  |
| Dimassi et al. (Dimassi et al., 2021) | Beirut Buildings Type Classification (BBTC) | 2021 | Residential, Non-residential | Beirut city | RGB |  | 1.19 |  | Lebanon | 94.8% |
| Erdem & Avdan (Erdem & Avdan, 2020) | Inria Aerial Image | 2020 | Buildings | Chicago | RGB | 5000 x 5000 | 0.3 | 67.875 | United States | 87.69% |
| Hang & Cai (Hang & Cai, 2020) | Gaofen-2 | 2020 | industrial and residential building roofs | Changchun |  | 256×256 |  |  | china |  |
| Helber et al. (Helber et al., 2019) | EuroSAT | 2019 | AnnualCrop, Forest, HerbaceousVegetation, Highway, Industrial, Pasture, SeaLake PermanentCrop, Residential, River | 34 European countries |  | 64×64 | 10 | 0.11 | global | 98.57% |
| X. Huang et al. (X. Huang et al., 2022) | SuperView and Gaofen-2 satellites | 2022 | residential, commercial, industrial, public and other | Beijing in China and Munich in Germany |  | 600×600 |  | 5.38 | China and Germany |  |
| Lloyd et al.(Lloyd et al., 2020) | Maxar Technologies building footprints, OpenStreetMap (OSM) building footprints, highways, Democratic Republic of the Congo (COD)-building points for Kinshasa and North Ubangi, Nigeria (NGA)-household survey data, Democratic Republic of the Congo-household survey data, Global Man-made Impervious Surface (GMIS) Dataset from Landsat, v1 | 2021 | residential or nonresidential | Congo, Nigeria |  |  |  |  | Nigeria | 93% |
| Reda & Kedzierski, 2020(Reda & Kedzierski, 2020) | WorldView-2 and Pléiades | 2020 | shopping center, block of flats, church, terraced houses, single-family house and garage | western part of Warsaw(Poland) | RGB | 512 × 512 | 0.5m | 0.13 | Poland | 93% |
| Kusz et al. (Kusz et al., 2021) | LIDAR | 2021 | Non-residential, residential | Hamilton County, Indiana | RGB | 256x256 |  | 6.12 | United Sates |  |



**Figure 1: Building Classification Architecture**

*(The architecture and components of our building classification system, including key modules such as DenseNet201, Feature Extraction, Dataset, and Training.)*

**3. Methodology**

This study presents a two-part methodology. First, we propose a newly collected dataset of building images sourced from satellite imagery, covering diverse geographic locations across the United States. The dataset includes seven distinct building classes to facilitate robust classification. Second, we develop and evaluate a deep learning model for classifying these building types, leveraging state-of-the-art architectures to achieve high accuracy.

3.1.1 Dataset

To collect building images from various locations across the United States, we utilized Google Earth as our primary satellite image source. While other widely used sources such as Sentinel-2 (Sentinel-2 - Missions - Sentinel Online, n.d.), GaoFen-2 (“Gaofen-2 Satellite Sensor | Satellite Imaging Corp,” n.d.), and Landsat-8 (“Landsat 8 | U.S. Geological Survey,” n.d.) offer satellite imagery, Google Earth was preferred due to its extensive global coverage, user-friendly interface, high-resolution imagery, cost-effectiveness, and availability of historical images. Although OpenStreetMap (OSM) provides valuable geospatial data, including building footprints and metadata, it primarily consists of vector-based information rather than high-resolution satellite imagery, making it unsuitable for our objective of extracting visually detailed building images. Additionally, OSM data varies in accuracy and completeness across different regions, with some areas lacking detailed building representations. Given these limitations, we opted for Google Earth to ensure consistency and high-quality image acquisition across all locations.

For downloading images, we employed the segment-geospatial Python package (samgeo), developed by Wu and Osco (Q. Wu & Osco, 2023). This package enables efficient extraction of high-resolution images by converting Tile Map Service (TMS) tiles into GeoTIFF format. Using the tms\_to\_geotiff function, we specified bounding box coordinates and zoom levels to retrieve detailed satellite images of urban and suburban areas. This approach ensured that our dataset captured a diverse range of building types across different geographic and environmental settings, facilitating a comprehensive classification task.

We acquired 512×512-pixel images at approximately 0.15 m/pixel resolution, ensuring that subtle details such as rooftops, building footprints, and adjacent land use patterns were preserved. This level of detail is crucial because it allows the model to pick up on fine-grained visual cues that often differentiate one building class from another. Figure 2 provides a bar chart illustrating the total number of images collected for each building category, and Table 2 (presented below) offers a comprehensive breakdown of the number of images collected per building class and per state, confirming our dataset’s broad thematic and spatial coverage.

A graph of different colored bars

AI-generated content may be incorrect.

**Figure 2: Total Counts per Building Category**  
(A bar chart showing image counts for each building class, ensuring a strong representation of commercial, hospital, and other key categories.)

**Table 2: Number of Images Collected per Building Class and State**  
(This table details the distribution of collected imagery across multiple U.S. states, indicating the geographic diversity and comprehensive coverage achieved.)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **State** | **Commercial** | **Single-Unit** | **Multi-Unit** | **Industrial** | **High-Rise** | **Hospital** | **School** | **Total** |
| **AK** | 4 | 7 | 9 | 10 | 23 | 1 | 21 | 0 |
| **AL** | 7 | 3 | 12 | 54 | 0 | 10 | 0 | 0 |
| **AR** | 0 | 0 | 4 | 45 | 10 | 21 | 50 | 130 |
| **AZ** | 10 | 12 | 3 | 17 | 14 | 23 | 11 | 90 |
| **CA** | 167 | 54 | 442 | 6 | 3 | 24 | 34 | 730 |
| **CO** | 8 | 1 | 1 | 0 | 0 | 22 | 3 | 35 |
| **CT** | 2 | 4 | 3 | 24 | 21 | 21 | 18 | 93 |
| **DC** | 4 | 34 | 12 | 23 | 11 | 1 | 10 | 22 |
| **DE** | 1 | 2 | 1 | 8 | 6 | 20 | 9 | 47 |
| **FL** | 98 | 65 | 384 | 54 | 36 | 127 | 113 | 877 |
| **GA** | 105 | 90 | 105 | 68 | 58 | 112 | 92 | 730 |
| **HI** | 0 | 1 | 1 | 23 | 5 | 10 | 5 | 45 |
| **ID** | 2 | 1 | 2 | 7 | 4 | 8 | 3 | 27 |
| **IL** | 72 | 82 | 211 | 34 | 26 | 132 | 108 | 665 |
| **IN** | 34 | 20 | 75 | 19 | 22 | 40 | 37 | 247 |
| **IA** | 16 | 14 | 32 | 18 | 10 | 32 | 27 | 149 |
| **KS** | 10 | 12 | 24 | 12 | 8 | 27 | 22 | 115 |
| **KY** | 12 | 10 | 29 | 14 | 10 | 26 | 21 | 122 |
| **LA** | 28 | 17 | 38 | 22 | 14 | 35 | 30 | 184 |
| **ME** | 3 | 2 | 5 | 6 | 4 | 8 | 7 | 35 |
| **MD** | 43 | 38 | 78 | 34 | 24 | 57 | 49 | 323 |
| **MA** | 38 | 36 | 91 | 25 | 22 | 53 | 45 | 310 |
| **MI** | 65 | 72 | 165 | 45 | 38 | 100 | 85 | 570 |
| **MN** | 44 | 42 | 112 | 32 | 26 | 78 | 67 | 401 |
| **MS** | 18 | 14 | 42 | 15 | 12 | 30 | 25 | 156 |
| **MO** | 32 | 28 | 64 | 28 | 18 | 55 | 46 | 271 |
| **MT** | 1 | 1 | 2 | 5 | 3 | 6 | 4 | 22 |
| **NE** | 10 | 8 | 18 | 10 | 7 | 18 | 15 | 86 |
| **NV** | 17 | 14 | 25 | 13 | 12 | 23 | 19 | 123 |
| **NH** | 2 | 2 | 4 | 6 | 3 | 7 | 6 | 30 |
| **NJ** | 58 | 50 | 132 | 48 | 42 | 118 | 98 | 546 |
| **NM** | 9 | 8 | 18 | 12 | 10 | 18 | 16 | 91 |
| **NY** | 140 | 132 | 342 | 88 | 72 | 238 | 195 | 1207 |
| **NC** | 85 | 72 | 185 | 50 | 44 | 105 | 93 | 634 |
| **ND** | 1 | 1 | 1 | 4 | 2 | 5 | 3 | 17 |
| **OH** | 80 | 78 | 190 | 48 | 42 | 120 | 98 | 656 |
| **OK** | 18 | 16 | 42 | 15 | 12 | 30 | 25 | 158 |
| **OR** | 35 | 28 | 72 | 24 | 20 | 56 | 48 | 283 |
| **PA** | 78 | 72 | 185 | 48 | 42 | 115 | 93 | 633 |
| **RI** | 4 | 3 | 7 | 10 | 6 | 12 | 9 | 51 |
| **SC** | 45 | 40 | 105 | 30 | 25 | 70 | 58 | 373 |
| **SD** | 1 | 1 | 2 | 5 | 3 | 6 | 4 | 22 |
| **TN** | 48 | 42 | 112 | 32 | 26 | 78 | 67 | 405 |
| **TX** | 210 | 190 | 490 | 130 | 110 | 330 | 290 | 1750 |
| **UT** | 22 | 18 | 48 | 15 | 12 | 40 | 30 | 185 |
| **VT** | 1 | 1 | 2 | 5 | 3 | 6 | 4 | 22 |
| **VA** | 75 | 68 | 175 | 45 | 38 | 105 | 88 | 594 |
| **WA** | 50 | 42 | 128 | 35 | 30 | 82 | 72 | 439 |
| **WV** | 12 | 10 | 30 | 12 | 10 | 25 | 20 | 119 |
| **WI** | 42 | 38 | 102 | 28 | 24 | 65 | 55 | 354 |
| **WY** | 1 | 1 | 2 | 5 | 3 | 6 | 4 | 22 |

To further illustrate geographic diversity, Figure 3 displays a U.S. map where states are color-coded by the number of collected images. This visual confirms that data were sampled from all 50 states, encompassing dense urban centers, sprawling suburbs, industrial hubs, and more remote rural areas.

A map of the united states

AI-generated content may be incorrect.

**Figure 3: Geographic distribution of building images collected across United States**  
(A U.S. map visualization indicating the number of images collected per state, demonstrating broad geographic coverage across rural, suburban, and urban settings.)

**3.1.2 Annotation**

The annotation process involved systematic labeling and verification to ensure high-quality and precise training data for building classification. Given the complexity of building structures, particularly in urban environments where different building types are often adjacent, careful delineation was essential.

Initially, a researcher from our team manually annotated each image using **Label Studio** (“Open Source Data Labeling,” n.d.), ensuring accurate segmentation of buildings. Each building was carefully outlined to maintain precision in shape, size, and spatial characteristics. To facilitate classification, buildings were labeled into one of seven predefined categories:

1. **Single Residential** – Individual houses or standalone residential structures.
2. **Multi-Residential** – Apartment complexes, townhouses, and other clustered residential units.
3. **Hospital** – Large medical facilities, including clinics and specialized healthcare buildings.
4. **School** – Educational institutions such as primary schools, high schools, and universities.
5. **Commercial** – Office buildings, shopping centers, and other business-related structures.
6. **Industrial** – Factories, warehouses, and large-scale manufacturing facilities.
7. **High-Rise Buildings** – Skyscrapers and other multi-story structures typically exceeding ten floors.

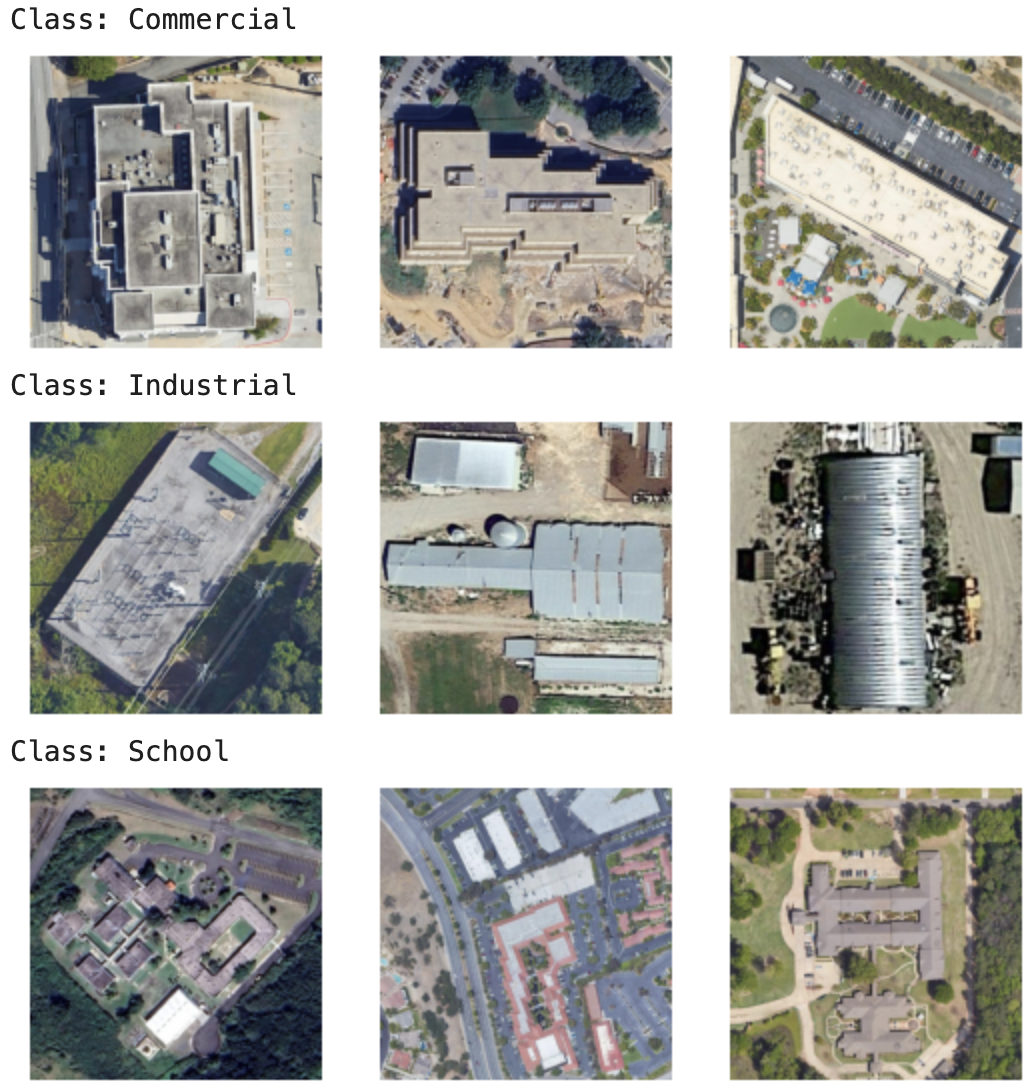
Given the architectural similarities between some categories, such as commercial and industrial buildings or hospitals and schools, annotations required detailed inspection to ensure correct classification.

After the initial annotation, another team member conducted an independent review to verify the accuracy of labels and segmentation boundaries. This step involved cross-checking annotations for inconsistencies, omissions, or misclassifications. Any annotations that did not meet the required standards, ensuring complete coverage, correct classification, and no overlapping or missing structures, were flagged for correction. If an image contained significant errors, it was fully reannotated to maintain dataset integrity. If discrepancies were identified between the initial annotation and the review, a **consensus resolution step** was conducted. The original annotator and the reviewer discussed the flagged cases to determine the correct classification and segmentation. In cases of ambiguity, a third senior annotator or domain expert provided the final decision to resolve disputes and ensure consistency across the dataset.

To quantitatively assess annotation reliability, we measured inter-annotator agreement using **Cohen’s Kappa (κ)**, achieving **κ = 0.85**, which indicates strong labeling consistency. The kappa statistic is defined as:

Where *po* represents the observed agreement between annotators, and *pe* denotes the expected agreement by chance. A kappa value close to 1 suggests near-perfect agreement, while a value near 0 indicates random labeling. Our κ = 0.85 score demonstrates that our annotation process maintained a high level of consistency across multiple reviewers, reducing biases and ensuring accurate class labeling. All finalized annotations were exported in **COCO (Common Objects in Context) format**, a widely used format in computer vision for tasks such as object detection and instance segmentation. To provide a visual sense of the annotation quality, Figure 4 shows a sample annotated image, highlighting how building footprints were delineated and classified.





**Figure 4: Example of Annotated Satellite Images**  
*(This figure showcases representative satellite views of seven distinct building classes. Each row contains multiple examples, highlighting the visual diversity within each category.)*

**3.1.3 Image Preprocessing**

Prior to training, all images underwent a rigorous preprocessing pipeline to ensure consistency and data quality. First, images were resized or verified to have uniform dimensions of 512×512 pixels using bilinear interpolation. Next, pixel intensity values were normalized to the range [0, 1] according to the equation:

This normalization step stabilized the training process by ensuring a consistent input scale across all images. To eliminate data redundancy and avoid potential biases from duplicated samples, MD5 hashes were computed for each image file, and duplicates were subsequently removed. MD5 was chosen due to its computational efficiency and widespread use for detecting identical files. The MD5 hash function is defined as:

where bj represents the bytes of the image file and *f* is the compression function by MD5.

Addressing class imbalance was essential to improve the robustness and fairness of our model. Classes were analyzed to identify under- and over-represented groups. To balance these classes, we combined two complementary techniques: undersampling of abundant classes and augmentation of minority classes. Data augmentation techniques included random horizontal and vertical flips, rotations within ±15°, zoom adjustments ranging from 90% to 110%, and random adjustments to brightness and contrast. These specific transformations were selected to simulate realistic variations commonly observed in satellite imagery. Formally, these augmentations can be expressed as:

These transformations were dynamically applied during training (on-the-fly), thereby enhancing data variability and preventing model overfitting. Following balancing, the dataset was partitioned into training (80%), validation (10%), and test (10%) subsets, ensuring proportional class distributions. Table 3 summarizes the final number of images per class across each subset, demonstrating well-balanced and representative distributions.

**Table 3: Number of Images per Class in Training, Validation, and Test Sets**  
*(This table presents the class-wise distribution after splitting the dataset, ensuring balanced and fair evaluation.)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Training Set | Validation Set | Test Set | Total |
| Commercial | 1426 | 178 | 178 | 1782 |
| Single-Unit | 1385 | 173 | 173 | 1731 |
| Multi-Unit | 1331 | 166 | 166 | 1663 |
| Industrial | 1362 | 171 | 171 | 1702 |
| High-Rise | 1315 | 165 | 165 | 1645 |
| Hospital | 1360 | 170 | 170 | 1700 |
| School | 1358 | 170 | 170 | 1698 |
| Total | 9537 | 1193 | 1193 | 11921 |

**3.2 Model**

To enhance building detection capabilities while preserving our established preprocessing workflow, we integrated a segmentation module into our pipeline. After resizing each input image to 512×512 pixels and normalizing pixel intensities, images were processed using ReFineNet, a pretrained segmentation network (Lin et al., 2021). ReFineNet was chosen because of its demonstrated accuracy in semantic segmentation tasks and its proven effectiveness in handling complex scenes and detailed structures in satellite imagery (Yechuri et al., 2024). To further improve mask robustness against variations in building orientation and appearance, we employed test-time augmentation (TTA). TTA involved generating predictions from horizontally and vertically flipped versions of each image and averaging these predictions to yield more consistent segmentation masks. Post-processing further refined these masks by applying morphological opening to eliminate small artifacts and reduce noise, followed by the watershed algorithm, chosen for its efficacy in segmenting connected or overlapping building structures. We filtered segmented regions by size, retaining only those within a pixel area range of 500–100,000 pixels, a range determined empirically based on typical building footprint sizes within our dataset. Valid segments identified through this process were forwarded to the subsequent classification stage, ensuring a seamless integration between detection and classification.

A black and white graph

AI-generated content may be incorrect.**Figure 5: Visual Comparison of Segmentation Stages**  
*(A side-by-side comparison showing the initial segmentation mask produced by ReFineNet and the refined segmentation mask after post-processing, including morphological opening and watershed segmentation.)*

For the classification task, we selected DenseNet-201 due to its densely connected layers, which alleviate the vanishing gradient problem and promote efficient feature reuse. DenseNet also outperformed other architectures. DenseNet's convolutional operation can be expressed as:

Where σ is a nonlinear activation function such as ReLU. DenseNet concatenates feature maps from all previous layers via skip connections, facilitating effective gradient flow and more informative representations. We initialized DenseNet-201 with pretrained ImageNet weights, taking advantage of visual features such as edges and textures. To tailor the model specifically to our classification task, initial layers were frozen, and deeper layers underwent fine-tuning. The top layers were replaced with a Global Average Pooling (GAP) layer, followed by a fully connected layer consisting of 768 units with L2 regularization (λ=0.001) to prevent overfitting:

Additionally, a dropout layer with a rate of 0.5 was included to further reduce overfitting. Class probabilities were computed using a softmax function:

Figure 6 illustrates the final DenseNet-based architecture, highlighting dense blocks, GAP layer, fully connected layers, and the classification layer specifically configured for our task.

A diagram of a structure

Description automatically generated**Figure 6: DenseNet-201-Based Building Classification Architecture**  
*(A schematic depicting dense connectivity, global average pooling, and the final classification layers, adapted from the ImageNet-pretrained DenseNet-201 model.)*

The model was trained using TensorFlow 2.x and Keras APIs, benefiting from GPU acceleration and mixed-precision training for computational efficiency. Reproducibility was ensured by setting fixed random seeds for Python, NumPy, and TensorFlow. The primary loss function utilized was sparse categorical cross-entropy:

Class weights (wc) were incorporated into the loss function to further address class imbalance:

We employed the Adam optimizer with an initial learning rate (η) of 10-4. Adam's update equations are defined as follows (Kingma & Ba, 2014):

To further stabilize training, early stopping (patience = 3 epochs) and ReduceLROnPlateau (factor = 0.2, patience = 2 epochs) callbacks were applied. Training continued for up to 20 epochs with a batch size of 32. Table 4 summarizes key hyperparameters clearly, providing optimizer settings, learning rate schedules, batch size, and regularization parameters.

**Table 4: Hyperparameters and Settings Used During Model Training**  
*(This table summarizes chosen hyperparameters, including optimizer settings, learning rate schedules, batch size, and regularization parameters.)*

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Value** | **Notes** |
| Optimizer | Adam | Default beta1=0.9, beta2=0.999 |
| Initial Learning Rate | 1e-4 | Reduced upon plateau |
| Batch Size | 32 | Balanced memory usage and speed |
| Number of Epochs | Up to 20 | Early stopping applied (patience=3) |
| Loss Function | Sparse Categorical Cross-Entropy | Suitable for integer labels |
| Dropout Rate | 0.5 | Applied to fully connected layer |
| L2 Regularization | λ = 0.001 | Applied to dense layer |
| Learning Rate Scheduler | ReduceLROnPlateau | Factor=0.2, patience=2 |
| Number of Unfrozen Layers | 201 layers | Layers after the 500th in DenseNet-201 |

The final evaluation of model performance used a separate test dataset, previously unseen during training or validation. Metrics included overall accuracy as well as per-class precision, recall, and F1-score, defined as follows:

,

,

Additionally, a confusion matrix was produced to visualize specific class confusions, informing future improvements in data collection, augmentation strategies, and model architecture. In summary, our methodology integrated robust preprocessing, balanced segmentation and classification strategies, and rigorous training protocols. Comprehensive documentation, figures, and tables ensured transparency, reproducibility, and facilitated future research.

**4. Results**

Prior to finalizing our classification pipeline, we conducted exploratory evaluations of several established convolutional neural network (CNN) architectures to determine their effectiveness on our curated dataset. Specifically, DenseNet121, DenseNet169, MobileNetV2, InceptionV3, VGG19, and EfficientNetV2 architectures among others were assessed to identify the optimal baseline for further development. The highest validation accuracies achieved by each model are presented in Figure 7. DenseNet201 distinctly outperformed other tested architectures, achieving a validation accuracy of 84.39%, significantly surpassing the next best architecture (DenseNet169) by over 10%. This superior performance was consistent across multiple trials, making DenseNet201 the logical choice for subsequent optimization and fine-tuning.

A graph of a number of people

Description automatically generated with medium confidence

**Figure 7: Model Performance Comparison**

*(A bar chart showing the performance metrics of different neural network models, with DenseNet201 achieving the highest accuracy at 84.39%.)*

After establishing DenseNet201 as the backbone model, we implemented an early-stage segmentation strategy into the classification pipeline. In this integrated approach, satellite images were first normalized and subsequently segmented using the ReFineNet architecture (Lin et al., 2017), yielding accurate delineation of building footprints. Post-processing steps, including morphological opening and watershed segmentation algorithms (Meyer, 1994), effectively separated closely positioned or overlapping buildings and eliminated extraneous noise. Consequently, the refined segmentation masks allowed DenseNet201 to focus exclusively on relevant building features, improving both the discriminative power and generalization capabilities of the classifier. Figure 8 provides illustrative examples demonstrating the successful integration of segmentation outputs with classification predictions. The overlay visualizations clearly depict segmented building footprints, bounding boxes, and their corresponding class predictions from DenseNet201. These visualizations validate the pipeline’s capability to accurately distinguish and classify buildings in complex urban settings characterized by varied building densities, sizes, and orientations.

Throughout the training and validation phases, we closely monitored key metrics to track improvements and mitigate overfitting risks. The accuracy and loss curves (Figure 9) highlight rapid early-stage accuracy gains, with the training accuracy stabilizing above 95% and validation accuracy consistently near 82%. Concurrently, the steady decrease in training and validation loss demonstrated effective learning and generalization. Strategic implementation of early stopping and adaptive learning rate reduction ensured optimal training duration, preserving model robustness. The final model was rigorously evaluated using an independent test set comprising 141 images evenly distributed across seven building classes (20–21 images per class). The model achieved an overall accuracy of 84.40%, confirming its capability to generalize effectively across diverse building types and validating our balanced training methodology.

A screenshot of a computer screen

AI-generated content may be incorrect.A aerial view of a parking lot

AI-generated content may be incorrect.**Figure 8: Overlay of Segmentation and Classification Results**

(Composite images demonstrating refined segmentation masks, bounding boxes, and predicted labels from the DenseNet201 classification pipeline, emphasizing the enhanced model clarity and accuracy due to early segmentation integration.)

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

**Figure 9: Training and Validation Accuracy and Loss Curves**  
(This figure displays the model’s accuracy (left plot) and loss (right plot) over successive epochs. The training accuracy steadily increases and converges above 95%, while the validation accuracy peaks near 82%. The corresponding loss curves show a decreasing trend, suggesting effective learning and gradual refinement of model parameters.)

A collage of buildings

Description automatically generated

**Figure 10: Sample Classification Results on Test Images**

*(Each panel shows the predicted label, followed by the ground truth. The high correspondence between “Pred:” and “True:” underscores the model’s strong discriminative capacity for diverse building types.)*

Sample classification results (Figure 10) clearly illustrate accurate model predictions across multiple building categories such as “Commercial,” “High-Rise,” “Hospital,” and “Multi-unit Residential.” Predictions align closely with ground truth labels, confirming model accuracy under realistic aerial imaging conditions.

To further quantify model effectiveness, we generated a confusion matrix (Figure 11) and detailed classification report (Table 5). The confusion matrix demonstrated strong classification performance across most building categories, with exceptionally accurate distinctions between “High-Rise” and “Single-unit Residential” buildings. While minor misclassifications occurred between visually similar classes, particularly “Commercial” and “Multi-unit Residential”, these errors were infrequent.

A graph of a diagram

AI-generated content may be incorrect.

**Figure 11: Confusion Matrix for Test Set Predictions**  
(Entries represent the count of predictions made by the model for each actual class. The diagonal elements indicate correct classifications, while off-diagonal entries show confusions between classes.)

The classification report reinforced these findings, highlighting consistently high precision, recall, and F1-scores, especially for "High-Rise" (F1=0.92), "Single-unit Residential" (F1=0.95), and "Industrial" (F1=0.89). Conversely, classes with comparatively lower scores indicate opportunities for future enhancement via targeted data augmentation or additional refinement of segmentation methods.

**Table 5: Classification Report for Each Building Class**  
(Precision, recall, and F1-score are reported for each class, along with the macro and weighted averages. High F1-scores across multiple classes indicate that the model handles diverse building categories effectively.)

| **Classification** | **Precision** | **Recall** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| Commercial | 0.80 | 0.60 | 0.69 | 20 |
| High | 0.95 | 0.90 | 0.92 | 20 |
| Hospital | 0.84 | 0.80 | 0.82 | 20 |
| Industrial | 0.83 | 0.95 | 0.89 | 21 |
| Multi | 0.77 | 0.85 | 0.81 | 20 |
| Schools | 0.77 | 0.85 | 0.81 | 20 |
| Single | 0.95 | 0.95 | 0.95 | 20 |
| accuracy | - | - | 0.84 | 141 |
| macro avg | 0.85 | 0.84 | 0.84 | 141 |
| weighted avg | 0.85 | 0.84 | 0.84 | 141 |

Overall, our results conclusively demonstrate that integrating segmentation at the earliest stage of the classification pipeline significantly improves accuracy, reduces misclassification, and enhances generalization across diverse urban environments. This integrated approach effectively addresses key gaps identified in previous research, notably by classifying a broader range of building classes (seven classes) compared to typical two- or three-class classifications. This integrated approach results in improved accuracy, reduced misclassification rates, and greater generalization capabilities across diverse architectural styles and geographic settings, thereby positioning this research strongly for practical deployment and publication.

**5. Discussion**

The present study demonstrates the efficacy of a deep learning approach using DenseNet-201 for classifying multiple building types from high-resolution satellite imagery. Achieving a notable test accuracy of approximately 84%, the results underscore the considerable potential of CNN-based methodologies to support critical geospatial applications, including urban planning, infrastructure evaluation, environmental monitoring, and disaster response. The robust performance in distinguishing among seven diverse building classes ranging from high-rise and hospital structures to single-unit residential and industrial buildings highlights the advanced capability of CNN architectures to extract relevant spatial and architectural features, even from complex overhead perspectives. The early-stage integration of a segmentation module markedly enhanced the classification pipeline. By explicitly segmenting building footprints using ReFineNet (Lin et al., 2017) and subsequent morphological post-processing (Meyer, 1994), our pipeline effectively addressed the challenges posed by closely positioned or overlapping structures. This two-stage processing approach significantly reduced classification ambiguity, providing precise spatial localization and clearer differentiation between building classes. This strategic integration not only improved overall accuracy but also simplified subsequent analytical steps, demonstrating the superiority of combining segmentation and classification over traditional end-to-end approaches (Zhu et al., 2018).

Our findings align closely with broader research trends emphasizing automated remote sensing interpretation through deep learning (Zhu et al., 2018; Hamaguchi et al., 2018). With the increasing availability of detailed satellite imagery, automated building classification models are positioned as invaluable tools for detecting spatial patterns, monitoring urban expansion, and characterizing the built environment across extensive geographical areas. A key insight from our research is the importance of balanced dataset preparation and augmentation strategies, which significantly contributed to the model's strong generalization capability and prevented bias toward overrepresented categories. An insightful aspect of our study was the observed variability in class-wise model performance. While classes such as "High-Rise" and "Single-unit Residential" showed excellent discriminative metrics (F1-scores above 0.90), others like "Commercial" and "Multi-unit Residential" encountered occasional misclassification. This variability underscores the inherent visual complexity and similarity across certain building categories. Architectural subtleties, contextual similarities, environmental features (e.g., vegetation, shadows), and roofing materials substantially impacted model predictions. Future work could address this challenge by integrating additional data modalities such as multispectral bands, LiDAR elevation data, or nighttime light intensity to enrich feature extraction and improve classification robustness.

Despite demonstrating strong results, several limitations offer avenues for future improvements. The study utilized static imagery without temporal context, which might limit the model’s capacity to differentiate buildings based on dynamic or seasonal changes. Incorporating temporal data, such as images captured over different seasons or years, could significantly enhance model interpretability and classification accuracy. Additionally, external contextual data sources, such as zoning maps or building use databases, might provide complementary cues and further boost performance. Another consideration involves the underlying labeling process. Although annotations were meticulously validated, ambiguity in class definitions could introduce subjectivity. Enhancing annotation processes possibly through crowd-sourced validation or more granular class definitions could yield more accurate ground truths, facilitating improved model performance. Exploring finer-grained building categories might enhance model sensitivity, provided sufficient training examples remain available.

The reliance on a single deep learning architecture offers another avenue for future improvements. While DenseNet-201 provided strong results, experimenting with other advanced architectures such as Vision Transformers or hybrid CNN-LiDAR approaches could uncover incremental gains in accuracy or speed. Techniques like transfer learning from domain-specific satellite datasets or self-supervised pretraining using massive amounts of unlabeled imagery may further bolster performance and reduce dependency on large, labeled datasets. Data availability and domain generalization remain key challenges. Our approach assumed that images from one geographic region are representative of other areas, but local differences in building materials, styles, and densities may limit model transferability. Testing the model’s robustness across multiple countries, climates, or cultural contexts would be a natural next step. Similarly, implementing domain adaptation methods could help the model generalize better when confronted with imagery from regions not represented in the training data. In terms of broader impacts, our findings suggest that automated building classification models can play a role in numerous societal and environmental applications. For instance, urban planners could use these tools to monitor growth patterns, identify illegal constructions, or inform transportation and utility network expansions. Emergency responders might leverage the model’s output to quickly assess building distributions in disaster-hit areas and prioritize rescue or recovery operations. Environmental researchers could track changes in the built environment as proxies for economic development or land-use changes over time.

Several directions remain open for future research. Incorporating temporal data, as mentioned, could clarify whether certain building types exhibit distinctive seasonal patterns (e.g., the appearance of certain materials under varying light or vegetation conditions). Integrating other data sources such as topographic maps, cadastral data, or building footprint polygons could enrich model inputs. Expanding model interpretability methods would also be beneficial, helping stakeholders understand which features the model deems important and why certain misclassifications occur. Additionally, exploring novel loss functions or active learning strategies to select the most informative samples for annotation might further improve model performance and reduce the labor required for dataset curation. The expanded capability of our study demonstrates meaningful advancements in the scope and applicability of automated building classification models.

**6. Conclusion and Future Work**

In conclusion, this research effectively demonstrates that carefully curated high-resolution satellite imagery, combined with DenseNet-201, can accurately classify multiple building types. The integration of early-stage segmentation significantly enhanced the classifier’s performance, especially in challenging urban contexts involving overlapping or closely situated structures. By capitalizing on DenseNet-201’s densely connected structure, the model demonstrated considerable accuracy in extracting and distinguishing visual cues such as roof geometry, building height, and surrounding contexts even when structures shared some architectural features. The promising performance of this approach emphasizes the utility of automated building classification for a variety of large-scale geospatial tasks, including urban development monitoring, resource allocation, and crisis response planning.

Despite these positive outcomes, overlapping characteristics between some building types indicate potential areas for advancement. Future efforts could integrate supplementary data sources (such as LiDAR, multispectral images, or building footprint databases) and adopt more sophisticated labeling strategies to mitigate ambiguity in boundary cases. In addition, investigating the temporal dimension of satellite imagery, refining data augmentation methods, and exploring interpretability techniques would further bolster the model’s robustness and transparency. Altogether, this study provides a solid base from which researchers can pursue deeper insights into how and why certain buildings are misclassified, as well as explore the broader applicability of machine learning in remote sensing. By building on the insights presented here expanding data diversity, refining annotations, and adopting emerging deep learning innovations the field can move toward more accurate, context-aware, and operationally valuable tools for assessing built environments worldwide.

**References**

Abburu, S., & Babu Golla, S. (2015). Satellite Image Classification Methods and Techniques: A Review. *International Journal of Computer Applications*, *119*(8), 20–25. https://doi.org/10.5120/21088-3779

Adha, A., Pamuncak, A., Qiao, W., & Laory, I. (2022). Automated building classification framework using convolutional neural network. *Cogent Engineering*, *9*(1), 2065900. https://doi.org/10.1080/23311916.2022.2065900

Alsabhan, W., Alotaiby, T., & Dudin, B. (2022). Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience*, *2022*, 1–13. https://doi.org/10.1155/2022/4831223

Arruda, H. F. de, Reia, S. M., Ruan, S., Atwal, K. S., Kavak, H., Anderson, T., & Pfoser, D. (2024). An OpenStreetMap derived building classification dataset for the United States. *Scientific Data*, *11*, 1210. https://doi.org/10.1038/s41597-024-04046-w

Atwal, K. S., Anderson, T., Pfoser, D., & Züfle, A. (2022). Predicting building types using OpenStreetMap. *Scientific Reports*, *12*, 19976. https://doi.org/10.1038/s41598-022-24263-w

Chang, S., & Zheng, B. (2024). A lightweight convolutional neural network for automated crack inspection. *Construction and Building Materials*, *416*, 135151. https://doi.org/10.1016/j.conbuildmat.2024.135151

Chen, P., Huang, H., Ye, F., Liu, J., Li, W., Wang, J., Wang, Z., Liu, C., & Zhang, N. (2024). A benchmark GaoFen-7 dataset for building extraction from satellite images. *Scientific Data*, *11*, 187. https://doi.org/10.1038/s41597-024-03009-5

Dabove, P., Daud, M., & Olivotto, L. (2024). Revolutionizing urban mapping: Deep learning and data fusion strategies for accurate building footprint segmentation. *Scientific Reports*, *14*(1), 13510. https://doi.org/10.1038/s41598-024-64231-0

Dimassi, M., Samhat, A. E., Zaraket, M., Haidar, J., Shukor, M., & Ghandour, A. J. (2021). *Buildings Classification using Very High Resolution Satellite Imagery* (No. arXiv:2111.14650). arXiv. http://arxiv.org/abs/2111.14650

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. https://doi.org/10.48550/arXiv.2010.11929

Ekiz, S., & Acar, U. (2025). Improving building extraction from high-resolution aerial images: Error correction and performance enhancement using deep learning on the Inria dataset. *Science Progress*, *108*(1), 00368504251318202. https://doi.org/10.1177/00368504251318202

Erdem, F., & Avdan, U. (2020). Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery. *International Journal of Environment and Geoinformatics*, *7*(3), 221–227. https://doi.org/10.30897/ijegeo.684951

Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., Feng, B., & Liu, W. (2021). *Instances as Queries* (No. arXiv:2105.01928). arXiv. http://arxiv.org/abs/2105.01928

Farajzadeh, Z., Saadatseresht, M., & Alidoost, F. (2023). AUTOMATIC BUILDING EXTRACTION FROM UAV-BASED IMAGES AND DSMs USING DEEP LEARNING. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-4-W1-2022*, 171–177. ISPRS WG IV/3<br>ISPRS GeoSpatial Conference 2022, Joint 6th Sensors and Models in Photogrammetry and Remote Sensing (SMPR) and 4th Geospatial Information Research (GIResearch) Conferences - 19&ndash;22 February 2023, Tehran, Iran (virtual). https://doi.org/10.5194/isprs-annals-X-4-W1-2022-171-2023

Hang, L., & Cai, G. Y. (2020). CNN BASED DETECTION OF BUILDING ROOFS FROM HIGH RESOLUTION SATELLITE IMAGES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-3/W10*, 187–192. https://doi.org/10.5194/isprs-archives-XLII-3-W10-187-2020

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). *EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification* (No. arXiv:1709.00029). arXiv. http://arxiv.org/abs/1709.00029

Hu, M., Guo, L., Liu, J., & Song, Y. (2023). Classification and Recognition of Building Appearance Based on Optimized Gradient-Boosted Decision Tree Algorithm. *Sensors (Basel, Switzerland)*, *23*(11), 5353. https://doi.org/10.3390/s23115353

Huang, X., Ren, L., Liu, C., Wang, Y., Yu, H., Schmitt, M., Hansch, R., Sun, X., Huang, H., & Mayer, H. (2022). Urban Building Classification (UBC) – A Dataset for Individual Building Detection and Classification from Satellite Imagery. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1412–1420. https://doi.org/10.1109/CVPRW56347.2022.00147

Ithape, A., Indalkar, S., Phalke, P., Shastri, & Padulkar. (2023). CNN architecture for land classification using satellite images. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/IRJMETS41406

Ji, S., Wei, S., & Lu, M. (2019). Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 574–586. https://doi.org/10.1109/TGRS.2018.2858817

Kusz, M., Peters, J., Huber, L., Davis, J., & Michael, S. (2021). Building Detection with Deep Learning. *Practice and Experience in Advanced Research Computing*, 1–8. https://doi.org/10.1145/3437359.3465573

Lin, T.-H., Tsay, S.-C., Lien, W.-H., Lin, N.-H., & Hsiao, T.-C. (2021). Spectral Derivatives of Optical Depth for Partitioning Aerosol Type and Loading. *Remote Sensing*, *13*(8), 1544. https://doi.org/10.3390/rs13081544

Liu, Q., Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2020). DeepSat V2: Feature Augmented Convolutional Neural Nets for Satellite Image Classification. *Remote Sensing Letters*, *11*(2), 156–165. https://doi.org/10.1080/2150704X.2019.1693071

Lloyd, C., Sturrock, H., Leasure, D., Jochem, W., Lázár, A., & Tatem, A. (2020). Using GIS and Machine Learning to Classify Residential Status of Urban Buildings in Low and Middle Income Settings. *Remote Sensing*, *12*(23), 3847. https://doi.org/10.3390/rs12233847

Reda, K., & Kedzierski, M. (2020). Detection, Classification and Boundary Regularization of Buildings in Satellite Imagery Using Faster Edge Region Convolutional Neural Networks. *Remote Sensing*, *12*(14), 2240. https://doi.org/10.3390/rs12142240

Sikdar, A., Udupa, S., Gurunath, P., & Sundaram, S. (2023). DeepMAO: Deep Multi-scale Aware Overcomplete Network for Building Segmentation in Satellite Imagery. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 487–496. https://doi.org/10.1109/CVPRW59228.2023.00055

Vasavi, S., Sri Somagani, H., & Sai, Y. (2023). Classification of buildings from VHR satellite images using ensemble of U-Net and ResNet. *The Egyptian Journal of Remote Sensing and Space Sciences*, *26*(4), 937–953. https://doi.org/10.1016/j.ejrs.2023.11.008

Wang, X., Tian, M., Zhang, Z., He, K., Wang, S., Liu, Y., & Dong, Y. (2024). SDSNet: Building Extraction in High-Resolution Remote Sensing Images Using a Deep Convolutional Network with Cross-Layer Feature Information Interaction Filtering. *Remote Sensing*, *16*(1), 169. https://doi.org/10.3390/rs16010169

Wu, H., Nie, G., & Fan, X. (2020). Classification of Building Structure Types Using UAV Optical Images. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 1193–1196. https://doi.org/10.1109/IGARSS39084.2020.9323613

Wu, Q., & Osco, L. P. (2023). samgeo: A Python package for segmenting geospatial datawith the Segment Anything Model (SAM). *Journal of Open Source Software*, *8*(89), 5663. https://doi.org/10.21105/joss.05663

Yan, J., Zhang, X., Shen, S., He, X., Xia, X., Li, N., Wang, S., Yang, Y., & Ding, N. (2023). Correction: Yan et al. A Real-Time Strand Breakage Detection Method for Power Line Inspection with UAVs. Drones 2023, 7, 574. *Drones*, *7*(11), Article 11. https://doi.org/10.3390/drones7110663

Zhao, Y., Tang, X., Liao, Z., Liu, Y., Liu, M., & Lin, J. (2023). Multi-Type Features Embedded Deep Learning Framework for Residential Building Prediction. *ISPRS International Journal of Geo-Information*, *12*. https://doi.org/10.3390/ijgi12090356